From biology to genes and back again: Gene discovery for monogenic forms of beta cell dysfunction in diabetes

Elisa De Franco

Institute of Biomedical and Clinical Science, University of Exeter Medical School, EX2 5DW, Exeter, UK

Corresponding author:

Elisa De Franco

Institute of Biomedical and Clinical Science,

Level 3, RILD Building

Barrack Road

EX2 5DW

Exeter

United Kingdom

E.De-Franco@Exeter.ac.uk

+44 132 408232

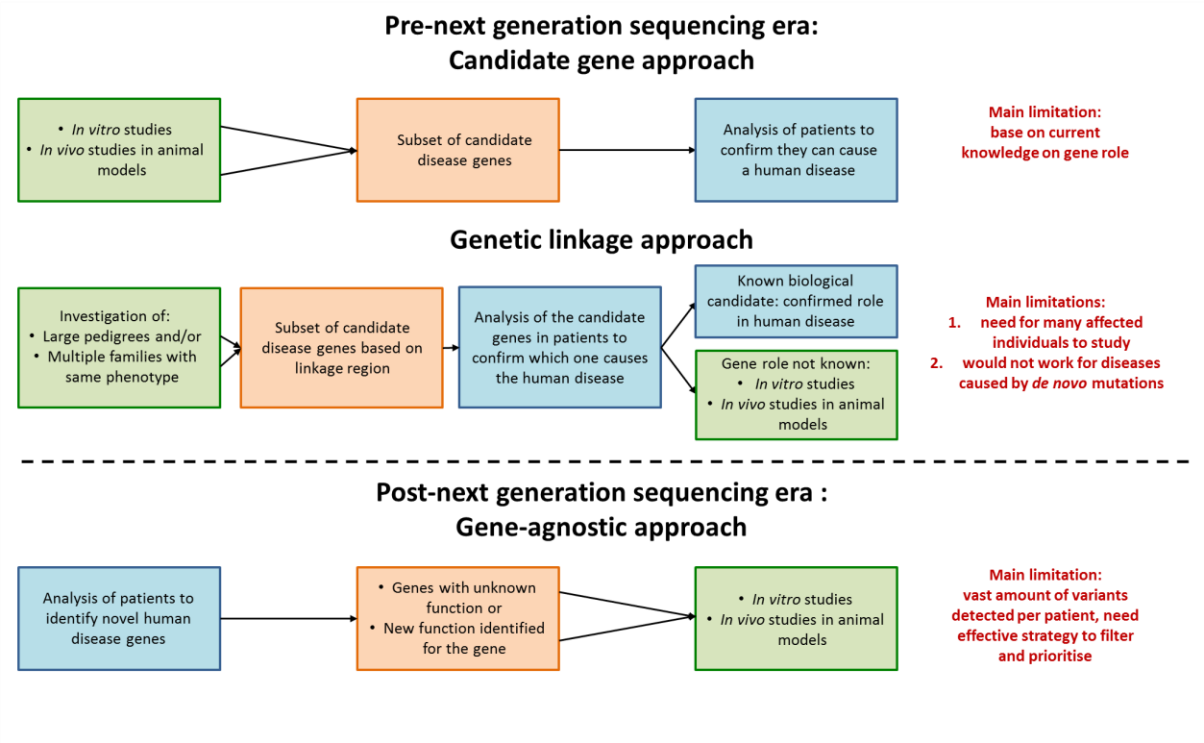
Abstract

This review focusses on gene discovery strategies used to identify monogenic forms of diabetes caused by reduced pancreatic beta cell number (due to destruction or defective development) or impaired beta cell function. Gene discovery efforts in monogenic diabetes have identified 36 genes so far. These genetic causes have been identified using four main approaches: linkage analysis, candidate gene sequencing and most recently, exome and genome sequencing.

The advent of next-generation sequencing has allowed researchers to move away from linkage analysis (relying on large pedigrees and/or multiple families with the same genetic condition) and candidate gene (relying on previous knowledge on the gene's role) strategies to use a gene agnostic approach, utilising genetic evidence (such as variant frequency, predicted variant effect on protein function, and predicted mode of inheritance) to identify the causative mutation. This approach led to the identification of 7 novel genetic causes of monogenic diabetes, 6 by exome sequencing and one by genome sequencing. In many of these cases, the disease-causing gene was not known to be important for beta cell function prior of the gene discovery study.

These novel findings highlight a new role for gene discovery studies in furthering our understanding of beta cell function and dysfunction in diabetes. Whilst many gene discovery studies in the past were led by knowledge in the field (through the candidate gene strategy) now they often lead the scientific advances in the field by identifying new important biological players to be further characterised by *in vitro* and *in vivo* studies.

Graphical abstract



1. Introduction to monogenic diabetes

Monogenic diabetes is caused by single gene mutations which most commonly results in beta cell dysfunction or destruction. There are two main subtypes, maturity-onset diabetes of the young (MODY) and neonatal diabetes. MODY is diagnosed during childhood or adulthood (>60% of cases diagnosed before 25 years) and accounts for ~3% of all cases of diabetes diagnosed under 30 years [1]. Neonatal diabetes is a much rarer condition (incidence ~1 case per 100,000 live births in European countries [2, 3]) and is diagnosed in the first 6 months of life. Monogenic forms of diabetes can also be one of the features of genetic syndromes involving multiple organs/tissues, for example maternally inherited diabetes and deafness (MIDD) caused by a mutation in the mitochondrial genome.

Targeted treatment for monogenic diabetes is one of the first examples of precision medicine where the genetic subtype determines the patient's treatment. Patients with MODY caused by mutations in the transcription factor genes *HNF1A* and *HNF4A* are sensitive to low dose sulphonylureas [4, 5] and those with *GCK* mutations usually do not require any pharmacological treatment. Most patients with neonatal diabetes caused by mutations in the genes encoding the K-ATP channel subunits Kir6.2 or SUR1 (*KCNJ11* or *ABCC8*) can be treated with high dose sulphonylureas [5, 6]. Stopping insulin improves quality of life [7], results in improved glycaemic control, reduces the risk of diabetic complications in later life and reduces healthcare costs [8]. Furthermore, improvements in cognitive function, motor skills and behaviour have been reported in patients with *KCNJ11/ABCC8* mutations causing neonatal diabetes with developmental delay [9-11].

The identification of monogenic forms of diabetes has highlighted how important obtaining a genetic diagnosis is for a patient's management. The clinical manifestations can be very similar between patients with different genetic subtypes, but the treatment and management options are often very different. The genetic diagnosis is therefore guiding patients' clinical management as it can result in improved treatment, defines the prognosis and the recurrence risk [12]. This is reflected in the current ISPAD guidelines for neonatal diabetes which advise immediate referral for genetic testing in patients diagnosed under the age of 6 months [12]. This has recently been reported to have resulted in neonatal diabetes patients now being referred for genetic testing less than 2 months after diagnosis with diabetes [13].

2. Gene discovery approaches in monogenic diabetes

In addition to having important implications for patients' clinical management, the identification of genes which, when disrupted, result in beta cell loss or dysfunction is very important for the diabetes research field as it can give fundamental insights into the pathogenesis of more common diabetes subtypes, such as type 1 and type 2. Furthermore, the identification of genes needed for beta cell formation during human embryonic development can provide important insights into the mechanisms regulating differentiation of stem cells into beta-cells. This knowledge could be crucial for current efforts aimed at generating functional beta cells from stem cells *in vitro*. Successful generation of beta cells *in vitro* is critical to overcome some of the main limitations currently being faced by islet transplantation approaches in patients with diabetes [13].

Before the introduction of next generation sequencing to monogenic disease research in 2010, DNA sequencing allowed only analysis of short fragments of DNA covering one portion of ~500bp of a gene at the time. The most widely used sequencing technique was the Sanger method developed in the 1970s. This approach was labour intensive and time consuming, allowing for only a limited number of genes to be sequenced in any given individual. This meant that gene discovery studies had to use effective prioritising strategies to select the genes to sequence. Traditionally, the most widely used (and successful) strategies for gene prioritisation have been the "candidate gene" and linkage analysis approaches. These approaches have resulted in the identification of the most common causes of neonatal diabetes (*KCNJ11* [14], *ABCC8* [15, 16], and *INS* [17, 18]) and MODY (*HNF1A* [19], *HNF4A* [20], and *GCK* [21]) and are further discussed in the following sections.

In the last decade the introduction of next-generation sequencing technologies has enabled researchers to perform DNA sequencing at several orders of magnitude greater throughput than was previously possible. This has resulted in a big change in the approaches needed for successful gene discovery. The possibility of sequencing all the genes in the human genome in a single experiment has shifted the main challenge from selecting the right gene to sequence to selecting the right variants to follow up.

Monogenic diabetes is clinically and genetically heterogeneous, with mutations in at least 36 genes reported so far (Table 1 and Figure 1). Discovery of these disease genes has resulted from linkage analysis, candidate gene sequencing and most recently, exome and genome sequencing. A list of monogenic diabetes genes, which strategy was used to identify them and a summary of the associated phenotypes are listed in Table 1. This review discusses the main gene discovery strategies and their most recent applications to identify genetic causes of monogenic diabetes, highlighting recent successes and challenges that remain to be addressed.

3. Monogenic diabetes found by linkage analysis

Investigation of large pedigrees with multiple affected family members has historically been a valuable approach to identify disease-causing genes. Traditionally this included the investigation of highly polymorphic microsatellite markers in affected and unaffected individuals to identify a genetic region in which all affected individuals shared the same allele(s). The bigger the pedigree, the more refined the region would be, allowing the investigators to select only a small number of genes to further investigate to identify the causative variant. In cases in which large pedigrees with multiple affected members were not available, this approach could also be used to investigate multiple small families thought to have the same genetic condition in order to identify a shared genomic region among the areas of linkage identified in the different families. This approach has been successfully used in the past to identify mutations in 18 genes causing monogenic forms of diabetes, including mutations in two pancreatic transcription factor genes causing pancreatic agenesis (*PDX1* [22] and *PTF1A* [23]), and more recently, to identify recessive mutations in the *TRMT10A* gene as a cause of young onset diabetes and primary microcephaly [24].

The main limitation to this approach is the requirement of large pedigrees and/or multiple pedigrees affected by the same genetic condition in order to be successful. This is often challenging when investigating a rare disease (like neonatal diabetes) or a disease with variable penetrance, such as MODY. Furthermore, whilst very useful for studying inherited conditions, a linkage analysis approach would not be effective when investigating a genetic condition caused by *de novo* mutations. However, linkage analysis remains a very powerful tool for gene discovery and it is still widely used in combination with next generation sequencing approaches to prioritise variants to follow up.

3.1 Recent identification of a novel syndromic form of monogenic diabetes through linkage analysis

The identification of the genetic aetiology of syndromic forms of monogenic diabetes is important to identify genes needed for the function of beta cells and other cell types. In 2013 Igoillo-Esteve *et al* reported the use of a combination of homozygosity mapping (a technique which identifies large stretches of homozygosity in the human genome) and linkage analysis to identify a homozygous nonsense mutation in the *TRMT10A* gene in two siblings born to related parents who had microcephaly and diabetes diagnosed in adolescence [24]. The authors went on to show that *TRMT10A* expression was high in both the adult and embryonic brain and in pancreatic islets.

The role of *TRMT10A* in glucose control was confirmed by the identification of a homozygous missense variant in three patients with microcephaly who presented with hyperinsulinaemic hypoglycaemia which progressed to diabetes in adolescence [25].

Prior to this report *TRMT10A* was not suspected to be important for beta cells and it is still unclear how absence of this gene results in beta cell dysfunction. Igoillo-Esteve and colleagues proposed a mechanism through which *TRMT10A* expression in beta cells is regulated by endoplasmic reticulum stress and absence of functional TRMT10A would therefore result in beta cell and neuron apoptosis [24].

4. Monogenic diabetes found by a gene candidate approach

The candidate gene strategy relies on the selection of a small group of genes which are suspected to be important for beta cell function and/or development based on previous studies conducted either *in vitro* or *in vivo* in animal models (most commonly mouse, frog, and zebrafish). This approach is based on the hypothesis that the mechanisms regulating beta cell function and development are conserved across species and therefore, if a gene is needed for beta cell development in mice, it is very likely to be needed for human beta cell development as well.

One of the biggest successes for this approach in monogenic diabetes has been the identification of activating pathogenic variants in the *KCNJ11* [14] and *ABCC8* [15, 16] genes as the most common cause of neonatal diabetes in non-consanguineous populations. This finding consequently led to improved treatment in these patients who can be effectively treated with sulphonylurea tablets rather than insulin injections.

Fifteen further monogenic diabetes genes have been identified using this strategy (Table 1), including most recently mutations in pancreatic transcription factors causing neonatal diabetes and adult onset diabetes as discussed below. This approach has therefore been very successful in the past, allowing the identification of some of the most common causes of monogenic diabetes. However as gene discovery efforts move towards rarer forms of the disease, using this strategy has become more challenging, mainly because of two limitations: 1) reliance on current knowledge and 2) prior assumption that the genes important for beta cell function in animal models are also fundamental for human beta cells.

4.1 Recent application of a candidate gene approach to find two novel genetic causes of neonatal diabetes resulting in defective beta cell development

The number of potential candidate genes for beta cell dysfunction is constantly increasing as more studies (including animal model phenotyping, genome wide association studies for type 1 and type 2 diabetes, and *in vitro* studies looking at expression and regulation of beta cells during differentiation) are published. For example, the International Mouse Phenotypic Consortium database [26] lists 534 genes associated with abnormal glucose homeostasis in mouse models. Prioritisation strategies are therefore needed to decide which genes to follow up. One recent example of how to refine the candidate gene approach has been reported by Flanagan and colleagues [27] who combined mapping of the homozygous regions with a candidate gene approach. In this study 29 transcription factor genes known to be important for pancreatic development in mice were systematically investigated in 147 patients with neonatal diabetes born to related parents. The authors hypothesised that these patients were likely to have a homozygous mutation causing their diabetes and used homozygosity mapping to investigate whether any of the 29 candidate pancreatic developmental factor genes were included in one of the large homozygous regions identified in the patients. This approach identified two patients with homozygous loss of function mutations in the *NKX2-2* gene and 2 patients with homozygous missense variants in *MNX1*.

The patients' clinical features were found to be very similar to the phenotypic characteristics described in mouse models in which the genes had been knocked out. The two patients with *NKX2-2* homozygous mutations had permanent neonatal diabetes and small corpus callosum. Consistent with this, the *Nkx2-2* knock-out mouse model was reported to develop diabetes soon after birth and to have reduced beta cell number and abnormal islet morphology [28]. The *Nkx2-2* knock-out mice also have neurological features including delayed oligodendrocyte differentiation and absence of hindbrain serotonergic neurons [29, 30].

The two patients with homozygous *MNX1* mutations reported by Flanagan *et al* had neonatal diabetes and severe intrauterine growth retardation. Only one of the two patients had additional extra-pancreatic features, whilst the other patient had neonatal diabetes and failure to thrive. This is clinically very different from the clinical features usually described in patients with Currarino syndrome which is caused by dominant loss of function *MNX1* mutations [31]. The *Mnx1* knock-out mouse model has intrauterine growth retardation and abnormal pancreatic development, with decreased beta cell number and abnormal islet morphology [32]. This is again very similar to the pancreatic phenotype observed in the two patients with neonatal diabetes.

This study had shown how successful a candidate gene approach can still be in the post next generation sequencing era, provided that the analysis of candidate genes is integrated with other prioritisation strategies, such as homozygosity mapping.

4.2 A candidate gene approach to identify *RFX6* dominant mutations as a cause of MODY

Mutations in the *HNF1A*, *HNF4A* and *GCK* genes are the most common genetic causes of isolated adult onset diabetes accounting for ~35% of patients with a MODY phenotype [1]. Identifying the genes that can cause isolated adult onset diabetes can give important insights into the pathogenesis of type 2 diabetes, by identifying factors and pathways needed for beta cell function. In order to try and identify novel genetic causes of MODY, Patel *et al* [33] used a next generation sequencing assay to analyse 29 genes known to cause neonatal diabetes, MODY and mitochondrial diabetes, lipodystrophy or other forms of syndromic diabetes in parallel in 38 patients who, based on their clinical features, were considered to be very likely to have a mutation in a single gene causing their diabetes. Using this approach the investigators identified two patients heterozygous for a nonsense *RFX6* variant. The frequency of heterozygous protein-truncating variants was then assessed in a larger cohort of 348 patients referred for MODY genetic testing and found to be significantly higher than in population datasets. This study led to the identification of *RFX6* mutations as the cause of MODY in 27 patients diagnosed at a median age of 32 years. Interestingly, the authors found that the penetrance of the *RFX6* heterozygous mutations was lower than for other MODY subtypes (namely *HNF1A* and *HNF4A*), and suggested that this is likely to be one of the reasons why gene discovery in adult onset diabetes is often very challenging.

RFX6 is crucially important for development of the human beta cells and biallelic mutations cause a syndromic form of neonatal diabetes which includes pancreatic hypoplasia/annular pancreas, gallbladder agenesis and intestinal atresia [34]. The results reported by Patel *et al* highlight RFX6's important role in adult beta cell physiology as well as development. This study also shows the value of candidate gene studies in large patient cohorts, allowing for replication of the initial genetic finding and overcoming the challenges presented by studying a heterogeneous disease with reduced penetrance.

5. Monogenic diabetes found by next generation sequencing - exome sequencing with a gene agnostic approach

The term next-generation sequencing collectively refers to the high throughput DNA sequencing technologies which are able to sequence a large amount of DNA sequences in a single experiment. The introduction of next-generation sequencing technologies to the market in 2005 and their subsequent improvement has resulted in the ability to sequence entire exomes and genomes quickly and at an affordable price for routine research and diagnostic tests. Next-generation sequencing approaches are now extensively used both for new disease genes discovery and for improving diagnostic genetic tests for known diseases.

The most widely used applications of next-generation sequencing are targeted sequencing of a panel of genes followed by exome sequencing with genome sequencing becoming more popular as prices fall. Exome sequencing allows simultaneous analysis of the ~2% of the human genome which encodes for proteins. About 80% of disease-causing mutations are predicted to be located in a protein-coding part of the genome (although this may be due to ascertainment bias as most studies only analyse the coding regions) [35], thus making exome sequencing an attractive strategy to investigate the genetic basis of monogenic diseases.

Typically, between 20,000 and 50,000 variants are identified per exome [36]. Filtering and prioritizing strategies are needed to reduce this number to a small subset of variants that are most likely to be pathogenic. The filtering steps applied to exome sequencing data account for qualitative requirements, predicted effect of the variant on the protein and whether the variant is known to be common in the general population (as reported in GnomAD [37] and internal databases when available). Generally, these steps leave 150–500 variants to be further assessed as potentially pathogenic [38]. This number is generally too large to allow follow-up of all the variants, and additional prioritisation strategies are needed. These strategies are generally based on the likely inheritance pattern of the disease (e.g. looking for recessive mutations in a linkage interval or *de novo* mutations in apparently sporadic disease). This strategy is often referred to as the 'gene agnostic' approach, as the filtering strategies used to prioritise the variants don't necessarily rely on prior knowledge of a gene's role, but on the predicted effect of the variant and the inheritance pattern.

More than 200 genes causing Mendelian disorders have been identified using exome sequencing, including six novel causes of monogenic diabetes [38-43].

5.1 Identification of new genetic causes of pancreatic agenesis by exome sequencing

Neonatal diabetes due to pancreatic agenesis is characterized by insulin dependent diabetes and pancreatic exocrine insufficiency requiring enzyme supplementation therapy [38]. This is an extremely rare condition which is most likely caused by a mutation in a single gene needed for pancreatic development. Identifying the genes that regulate pancreatic development in humans can give important insights into the factors needed to make functional beta cells and has the potential to be translated into optimisation of *in vitro* protocols to differentiate stem cells into beta cells.

Until 2012, only recessive mutations in two pancreatic developmental factors, *PDX1* and *PTF1A*, had been identified to cause pancreatic agenesis in humans through a combination of linkage and candidate gene studies. Mutations in *PDX1* had been described in 4 cases with isolated agenesis of the pancreas [22, 44, 45]. Mutations in *PTF1A* had been reported in 4 families in which affected individuals had both pancreatic and cerebellar agenesis [23, 46, 47].

In 2012 Lango Allen and colleagues investigated 27 patients diagnosed with pancreatic agenesis [38] and reported that most patients with syndromic pancreatic agenesis were born to unaffected unrelated parents. This suggested that the mutation causing syndromic pancreatic agenesis was most likely sporadic. To investigate this hypothesis the authors performed exome sequencing of two unrelated patients with pancreatic agenesis and congenital heart defects and their unaffected, unrelated parents with the aim of investigating variants present in the affected patient but not inherited from either parent. After exclusion of common variants, only one *de novo* variant was confirmed in each patient. Both variants, a missense and a frameshift deletion, affected the coding region of the developmental factor gene *GATA6*. Lango Allen and colleagues then sequenced *GATA6* in 25 additional patients with pancreatic agenesis and identified mutations in 13 additional cases [38]. The authors concluded that heterozygous mutations in *GATA6* are a common cause of pancreatic agenesis.

GATA6 is a transcription factor involved in early embryonic development of multiple organs, including the pancreas [48]. Interestingly, previous studies on mouse models were not suggestive of a role of Gata6 in pancreatic development in rodents [48, 49] and therefore investigation of *GATA6* in patients with neonatal diabetes had not been considered before. In this case exome sequencing led to the identification of a novel disease gene and gave unexpected insights into human pancreatic development.

The identification of mutations in *GATA6* as a major cause of syndromic pancreatic agenesis in humans highlighted the potential of the gene agnostic approach, demonstrating the existence of

fundamental differences in the genes regulating pancreatic development in mouse and human, an important limitation of the 'candidate-gene' approach.

The difference between development of the pancreas in mice and human was further highlighted by the recent report of a specific *CNOT1* mutation, p.(Arg535Cys), as causing a novel syndrome of pancreatic agenesis and holoprosencephaly [40]. In this study, De Franco and colleagues performed exome sequencing for 9 patients with pancreatic agenesis and their unaffected, unrelated parents (available in 7 cases). The same novel, heterozygous *CNOT1* missense mutation was found in 3 patients and confirmed to have arisen *de novo* in 2 of them (the maternal sample was not available for the third patient). Two of the patients were diagnosed with pancreatic agenesis and partial holoprosencephaly (a neurodevelopmental condition resulting from failure of the brain to separate into two hemispheres). The third patient had pancreatic agenesis and dysmorphic facial features suggestive of possible holoprosencephaly, but this could not be confirmed as MRI was declined by the parents. The identification of the same *de novo* variant in 3 patients with a very similar phenotype, led the authors to hypothesise that a mutation-specific mechanism rather than haploinsufficiency was responsible for the phenotype.

To test this hypothesis, the authors used CRISPR/Cas9 genome editing to generate a mouse model harbouring the *Cnot1* p.(Arg535Cys) variant. Interestingly, mice heterozygotes for the *Cnot1* mutation, which theoretically should have mirrored the human disease, had normal glucose tolerance and no gross abnormalities. However, homozygosity for the mutation was embryonically lethal and analysis of the embryos at e14.5 showed a range of brain defects and a markedly small dorsal pancreas compared to wild type and heterozygotes littermates [40].

A second study investigating the genetic causes of holoprosencephaly reported detection of the *CNOT1* p.(Arg535Cys) mutation in a patient with holoprosencephaly and neonatal diabetes (the second patient reported did not have diabetes), confirming the role of this mutation in pancreatic and brain development [50].

CNOT1 is a repressor of transcription known to act both as an independent factor and as the scaffold protein of the CCR4-NOT complex [51]. Before the recent identification of the p.(Arg535Cys) mutation in patients with syndromic pancreatic agenesis, it had never been thought to have a specific role in brain and pancreatic development. I*n vitro* studies had however suggested that CNOT1 is needed for maintaining human and mouse embryonic stem cell pluripotency through inhibition of early endodermal differentiating factors [52], including the known pancreatic developmental transcription factors GATA4 and GATA6. De Franco and colleagues have therefore

proposed a possible mechanism through which the p.(Arg535Cys) mutation causes pancreatic agenesis [40]. They suggested that the mutation could result in CNOT1 maintaining its inhibition on the GATA factors which in turn would result in continued expression of the SHH factor (which needs to be switched off for pancreatic development) and affected pancreatic development.

This finding highlights a possible new mechanism disrupting pancreatic development by affecting the very early stages of embryonic stem cell differentiation and reiterates the importance of gene-agnostic approaches to gain new biological insights into human pancreatic development.

5.2 Identification of new genetic causes of monogenic autoimmunity by exome sequencing

In some cases diabetes diagnosed before 6 months can be caused by mutations in a single gene causing severe early-onset autoimmunity leading to beta cell destruction. The most common of these conditions is IPEX syndrome (Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked), caused by mutations in the *FOXP3* gene [53]. Identification of the genes causing these conditions can give important insights into the mechanisms involved in the pathogenesis of more common autoimmune diseases, such as type 1 diabetes.

In order to identify novel genes causing early-onset syndromic autoimmune disease, Flanagan and colleagues [41] performed exome sequencing of a proband/parents trio for a patient diagnosed with diabetes at 2 weeks and additional early-onset autoimmune features (autoimmune hypothyroidism diagnosed at 3 years and celiac disease diagnosed at 17 months). A single *de novo* mutation in the transcription factor gene *STAT3* was identified. Sequencing of *STAT3* in 63 additional patients identified 3 mutations in 4 individuals.

Functional studies on the mutated STAT3 protein showed that the variants identified in patients with the early-onset polyautoimmunity phenotype were all activating mutations, whilst inactivating *STAT3* mutations had been previously reported to cause the immunodeficiency disease, Hyper IgE syndrome [54]. The authors propose a mechanism in which *STAT3* activating mutations lead to early autoimmunity by impairing the development of regulatory T cells [41].

The identification of *STAT3* mutations in patients with syndromic monogenic autoimmune disease has opened the possibility of personalised therapies to treat at least some of the most debilitating features of the condition in these patients. One example was reported by Milner *et al* [55] who showed significant improvement of the polyarthritis and skin tightening in a patient with a *STAT3* activating mutation when they were treated with an anti-IL6R monoclonal antibody.

Another genetic cause of monogenic autoimmunity including diabetes was recently reported by Johnson *et al* [42]. This study described the use of exome sequencing in a patient with neonatal diabetes and additional autoimmune features to identify compound heterozygous loss of function mutations in the *LRBA* gene. The authors conducted replications studies in 169 patients diagnosed with diabetes before the age of 12 months to identify a further 9 patients from 8 families with biallelic *LRBA* mutations and early onset diabetes (with age at diagnosis ranging from 6 weeks to 15 months). Most patients had additional autoimmune features, including haematological disorders, enteropathy, hypothyroidism and recurrent infections. Interestingly, one patient had isolated permanent neonatal diabetes when last assessed at the age of 2 years, suggesting that diabetes can be the presenting feature of this phenotypically heterogeneous disorder.

Biallelic mutations in *LRBA* had been previously identified in 2012 to cause common variable immunodeficiency with autoimmunity [56] however diabetes was considered to be a rare feature of the disease and, when present, was diagnosed outside the neonatal period. Johnson *et al* have reported that *LRBA* mutations were picked up in a high proportion of patients born to related parents and diagnosed with diabetes between 6 and 12 months of age, recommending inclusion of this gene in current genetic testing strategies for neonatal and early onset diabetes [42].

Early detection of an *LRBA* mutation in patients is crucially important as it can guide treatment decisions. When Loper-Herrera *et al* first reported the identification of mutations in *LRBA* as a cause of common variable immunodeficiency with autoimmunity using linkage analysis the gene function was completely unknown. Following this discovery, Lo *et al* hypothesised that LRBA could be controlling the expression of one of the master-regulators of immunity, CTLA4 [57]. This was confirmed by the marked clinical improvement of most symptoms in 9 patients with recessive *LRBA* mutations who were treated with Abatacept, a fusion drug replacing CTLA4 already used for treatment of rheumatoid arthritis. This is an important example of how the identification of mutations in *LRBA* as a cause of monogenic autoimmunity led to defining the role of the gene and developing a targeted therapy for the patients with this rare monogenic condition.

In the cases discussed in this section the identification of two novel causes of early onset diabetes using exome sequencing gave important insights into the complex mechanisms leading to autoimmunity and has contributed to development of better therapeutic options for patients.

5.3 Identification of a syndromic form of adult onset diabetes by exome sequencing

Exome sequencing can be extremely successful in identifying the genetic aetiology for clinically well-defined entities for which the previous approaches had failed. This was indeed the approach used by Cordeddu *et al* to identify the genetic cause of Primrose syndrome [58], a neurodevelopmental disorder often associated with a microdeletion of 5 genes at the 3q13 region. Exome analysis of 4 affected individual and their unrelated, unaffected parents identified *de novo* heterozygous mutations in the zinc finger transcription factors *ZBTB20* in all four patients. Replication studies identified heterozygous *ZBTB20* mutations in 4 further individuals with a suspected diagnosis of Primrose syndrome. When clinically assessed, it was found that 4/7 patients with *ZBTB20 de novo* mutations had diabetes and 3/7 had abnormal glucose tolerance (data was not available on the 8th individual).

Although the mechanism through which *ZBTB20* mutations cause diabetes in patients with Primrose syndrome is unclear, mouse studies had previously shown that *Zbtb20* is highly expressed in beta cells. The function of this transcription factor in the mouse beta cell was further confirmed by the characterisation of a beta cell-specific *Zbtb20* knock out mouse model which was found to develop hyperglycaemia through severely impaired glucose-stimulated insulin secretion [59].

The identification of heterozygous *de novo ZBTB20* mutations in patients with Primrose syndrome and diabetes confirms that this gene is fundamentally important for human as well as mice beta cells, highlighting a pathway which could be relevant to the pathogenesis of type 2 diabetes.

5.4 Identification of a novel WFS1-related disease by exome sequencing

Biallelic mutations in *WFS1* cause Wolfram syndrome [60], a degenerative condition characterised by early-onset diabetes (median age at onset 6 years, range 1–32 years [39]), which is often the presenting feature, followed by the development of optic atrophy, diabetes insipidus and deafness. Heterozygous dominantly-acting *WFS1* variants have been reported to cause 'milder' or 'incomplete' forms of Wolfram syndrome, including isolated deafness [61, 62], isolated nuclear cataracts [63], deafness and optic atrophy [64-66]. In 2013 Bonnycastle *et al* used a combination of exome sequencing and linkage analysis to identify the genetic cause of diabetes (diagnosed between 18 and 51 years of age) in a multigenerational Finnish family [67]. They identified a novel *WFS1* heterozygous variant, p.(Trp314Arg), which co-segregated with diabetes in the family. *In vitro* functional studies showed that this variant was likely to impair the WFS1 protein's ability to suppress endoplasmic reticulum stress after its induction.

More recently, De Franco and colleagues reported the identification of a *de novo* heterozygous *WFS1* variant, p.(Glu809Lys), in a patient with neonatal diabetes, congenital cataracts, sensorineural deafness, hypotonia, and dysmorphic features [39]. Replication studies identified *de novo* heterozygous missense variants in *WFS1* in 4 further patients, including 2 additional patients harbouring the same *WFS1* p.(Glu809Lys) mutation as the index patient. The clinical features were remarkably similar among the 5 patients, with all of them having low birth weight, early onset diabetes (diagnosed between 13 and 50 weeks of age) and deafness. Four of the five patients also had congenital cataracts and hypotonia. The authors noted that the phenotype in these 5 patients was very different from classical Wolfram syndrome, leading to the hypothesis that the mutations identified have a dominant negative effect. Functional *in vitro* studies confirmed this hypothesis, showing that the mutations resulted in a misfolded protein which causes severe endoplasmic reticulum stress as well as resulting in a WFS1 protein unable to suppress the endoplasmic reticulum stress response.

This study expanded the clinical spectrum associated with mutations in *WFS1*, defining a novel syndromic form of neonatal diabetes. These findings give important insights on how the WFS1 protein is fundamentally important for beta cell survival and highlight the potential of gene agnostic approaches in identifying new genetic mechanisms involving known disease genes.

5.5 Identification of *EIF2S3* mutations as causing an X-linked syndrome which can include early onset diabetes

Exome sequencing of proband/parents trios is very effective to identify *de novo* and homozygous/compound heterozygous variants. It is however more challenging to identify pathogenic variants that cause disease through an X-linked recessive mode of inheritance where affected males often inherit the variant from unaffected (or mildly affected) mothers.

Skopkova and colleagues successfully identified *EIF2S3* variants as the cause of X-linked MEHMO syndrome (a complex disease including mental retardation, epileptic seizures, hypogonadism and hypogenitalism, microcephaly, obesity and, in some cases, early onset diabetes) by performing exome sequencing on 4 affected individuals with similar clinical features [43]. Since all the affected were males, the authors focussed their analysis on rare, predicted deleterious variants on the X chromosome. All four patients were found to be hemizygous for rare *EIF2S3* variants. Functional studies on patients' fibroblasts and in yeast cells confirmed that the mutations were severely affecting the function of the protein encoded by EIF2S3.

EIF2S3 encodes for the translation initiation factor eIF2γ which is important for regulation of protein translation both under basal conditions and when there is endoplasmic reticulum stress. The results of this gene discovery study confirm the importance of this gene for many cell types, including beta cells.

6. Genome sequencing

The cost of genome sequencing has been steadily falling in recent years, leading many researchers to prefer this more comprehensive approach to exome sequencing. Genome sequencing allows ~~the~~ analysis of almost the entire genomic sequence (~98% [68]), without prior selection of specific regions. Each genome sequenced produces about 200 Gb of data with 3-4 million single nucleotide variants expected to be detected in each individual.

Genome sequencing presents some important technical advantages compared to exome sequencing: it is more sensitive and accurate for detecting structural variation (such as insertions, deletions, and translocations), it allows more even coverage than exome sequencing allowing more accurate variant detection, and can identify variants in intronic and intergenic regions. However, although genome sequencing is considered the most comprehensive strategy to date [69], the technology still presents some important limitations, mainly due to the use of short (~150-300 bp) fragments sequencing. This is currently a necessary requirement to obtain high sequence quality, but it also means that complex genomic regions (for example regions with very high GC content or highly repetitive elements), are impossible to map and therefore to analyse. This can result in genetic diagnosis being missed. The advent of technologies allowing sequencing of long (usually >10 kilobase) reads from single DNA molecules, such as the Oxford Nanopore [70] and PacBio [71] technologies, is a promising avenue to overcome these limitations. These technologies have been used successfully to investigate complex rearrangements and tandem repeats in know disease genes. However their routine use for human genome sequencing is currently hampered by the high error rate (~13%) and the high cost/throughput ratio, which is however constantly decreasing.

6.1 Identification of *PTF1A* non-coding variants causing pancreatic agenesis using genome sequencing

The main obstacles to the widespread use of (short read) genome sequencing for gene discovery have been the the enormous amount of data produced, resulting in challenging data analysis. Most of the studies reporting the use of genome sequencing so far have limited the initial variants analysis

to the coding part of the genome and have proceeded to the investigation of the non-coding variants just when a causative variant could not be identified in the exome.

So far, the most successful application of whole genome sequencing in monogenic diabetes has been the identification of mutations in a previously unrecognised regulatory element of the *PTF1A* gene [72] in patients with isolated pancreatic agenesis.

Recessive mutations in the gene encoding for the transcription factor *PTF1A* are a known cause of pancreatic and cerebellar agenesis. In 2014 Weedon and colleagues reported the use of linkage and genome sequencing to identify the genetic cause of isolated pancreatic agenesis in a cohort of patients born to related parents [72]. The authors studied 3 consanguineous pedigrees which included multiple affected individuals, suggesting a recessive pattern of inheritance. Linkage analysis in the 3 families highlighted a single shared locus on chromosome 10, including the *PTF1A* gene. However no coding mutation segregating with the disease was identified.

Genome sequencing was subsequently performed in 2 patients. Initial analysis of the coding variants failed to identify the genetic cause. The authors then concentrated on variants affecting regulatory elements involved in early pancreatic development identified by epigenomic mapping in pancreatic progenitor cells. A single shared homozygous variant located in a highly conserved region ~25kb downstream of *PTF1A* was identified. Sequencing analysis of the ~500bp-long putative regulatory element in 19 additional probands with pancreatic agenesis identified a mutation in 8 individuals.

Functional studies confirmed that the regulatory element was a previously unrecognised *PTF1A* enhancer which is active only during pancreatic development [72]. The authors suggested that this is likely to explain why patients with mutations in the *PTF1A* distal enhancer do not present with the severe cerebellar phenotype caused by the majority of *PTF1A* coding mutations [23, 46, 47].

This study is an important example of the application of genome sequencing to identify pathogenic non-coding variants and uncovers the role of a previously unsuspected regulatory element needed for normal pancreatic development in humans.

7. Challenges and opportunities

Following the initial successes of candidate gene and linkage analysis approaches, the introduction of next generation sequencing technologies has resulted in the identification of 6 novel genetic causes of neonatal-early onset diabetes and 1 novel genetic cause of syndromic adult onset diabetes though beta cell dysfunction in the last 7 years.

The identification of 6 novel genetic causes of neonatal diabetes has increased the proportion of patients in whom a genetic aetiology can be successfully identified to over 82% [12]. This proportion is even higher in patients with pancreatic agenesis for whom mutations in the known genes account for 97% of cases [40]. This is in sharp contrast with the success rate of gene discovery in MODY patients, for whom the pick-up rate is 35-45% [1]. One of the main challenges that explain why gene discovery has so far been less successful in MODY cohorts is patient selection for gene discovery studies. Whilst for neonatal diabetes the age at diagnosis cut-off of 6 months [2, 73] enriches patient cohorts for individuals who are likely to have a mutation in a single gene, it remains extremely challenging to distinguish the rare patients with adult onset monogenic diabetes from the much more common type 1 and type 2 diabetes using clinical features alone. A powerful tool that has the potential to overcome this obstacle is the use of genetic risk scores to identify patients who are genetically at high risk of developing type 1 [74-76] and type 2 diabetes [77, 78]. Pre-screening patients using genetic risk scores before gene discovery studies could greatly help researchers to select patients who are most likely to have a monogenic cause for their disease.

Another important challenge for gene discovery studies is the rarity of some genetic conditions, making replication of the genetic finding in three unrelated families an extremely challenging and lengthy process. Some international platforms such as GeneMatcher [79] are addressing this issue by allowing scientist from all over the world to input a candidate gene of interest in their freely available web-app and putting them in contact with other centres who have also submitted an entry for the same gene. These efforts are aimed at making the genetic replication process for rare genetic diseases easier and more efficient by encouraging scientists to share their data and building new international collaborations.

The vast majority of the gene discovery studies published so far have only analysed the ~2% of genomic sequence which is known to encode for proteins (the exome). As described above, this approach has been very successful in many cases, but there are still >10% of patients with neonatal diabetes and ~50% of patients with adult onset diabetes without a genetic diagnosis. Whilst some of these patients might have a very rare mutation in a novel gene which has not been replicated in a second family yet, it is likely that at least some of them have a mutation in the ~98% of the genome that was not initially investigated. Non-coding variants could affect splicing or be located in important regulatory regions (promoters or enhancers) which can be tissue and developmental stage-specific (as in the case of the *PTF1A* enhancer mutations in patients with pancreatic agenesis [72] discussed in the previous sections). The next challenge for gene discovery will be to develop strategies to prioritise, filter and interpret the large amount of non-coding variants identified by

genome sequencing to identify disease-causing mutations. The identification of non-coding (in particular regulatory) mutations causing diabetes can shed important insights onto the genomic regions regulating beta cell function and potentially offer insights into the mechanisms of other types of diabetes such as type 1 and type 2.

8. Perspective – a new role for gene discovery

The recent successes in using gene agnostic approaches to identify genetic causes of monogenic diabetes have highlighted a new role for gene discovery in science. Before, when the candidate gene approach was the most commonly used strategy, the identification of a genetic mutation causing diabetes in humans was often confirmatory of the role of a candidate gene in human beta cells.

Now, the use of a gene agnostic approach often results in the identification of disease-causing mutations in genes that either have no known function or had not been previously thought to be important for beta cells. This means that gene discovery is often becoming the starting point from which to perform further *in vitro* and *in vivo* studies to accurately define the role of the gene and the mechanism through which it is causing the disease. Gene discovery is therefore opening new avenues of research, highlighting novel biological mechanisms which are often specific to humans and would be very hard (and in some cases impossible) to identify using animal models. If gene discovery was previous led by the known science, today it is actively leading to new biological discoveries.

Acknowledgements

The author would like to thank Dr Thomas Laver and Dr Sarah Flanagan for critical revision of the manuscript. Elisa De Franco is the recipient of an EFSD Rising Star Fellowship.

Table 1. Monogenic forms of diabetes. This table lists monogenic form of diabetes caused by defective pancreatic beta cell function, development or destruction that would be classified as diagnostic-grade according to the criteria used by PanelApp (https://panelapp.genomicsengland.co.uk/#!Guidelines) and ClinGen [80]. This is not a comprehensive list of syndromes which can include diabetes. In addition, methylation defects that result in the overexpression of the paternally inherited genes *PLAGL1* and/or *HYMAI* genes are the most common cause of transient neonatal diabetes although the exact molecular mechanism is unknown [81]

Gene	Inheritance	Phenotype	Method of discovery	Reference
ABCC8	Dominant or recessive	Transient or permanent neonatal diabetes, DEND syndrome or MODY	Candidate gene	[15, 16, 82, 83]
CEL	Dominant	Diabetes and pancreatic exocrine dysfunction	Linkage	[84]
CISD2	Recessive	Wolfram Syndrome 2	Linkage	[85]
CNOT1	Dominant (*de novo*)	Pancreatic agenesis and holoprosencephaly	Exome sequencing	[40, 50]
DCAF17	Recessive	Woodhouse-Sakati syndrome	Linkage	[86, 87]
EIF2AK3	Recessive	Wolcott-Rallison syndrome	Linkage	[88]
EIF2S3	X-Linked Recessive	MEHMO syndrome	Exome sequencing	[43, 89, 90]
FOXP3	X-linked recessive	Immune dysregulation, polyendocrinopathy, enteropathy (IPEX) syndrome	Candidate gene/ Linkage	[53, 91, 92]
GATA4	Dominant	Permanent/Early onset diabetes with cardiac defects	Candidate gene	[93, 94]

GATA6	Dominant (often *de novo*)	Permanent or transient neonatal diabetes with cardiac, biliary or gut malformations and/or other endocrine abnormalities or MODY	Exome sequencing	[38, 95]
GCK	Dominant or recessive	Mild fasting hyperglycaemia (dominant) or permanent neonatal diabetes (recessive)	Linkage/Candidate gene	[21, 96]
GLIS3	Recessive	Permanent neonatal diabetes with congenital hypothyroidism	Linkage	[97]
HNF1A	Dominant	MODY	Linkage	[19]
HNF4A	Dominant	MODY	Linkage	[20]
HNF1B	Dominant (often *de novo*)	Renal cysts and diabetes (RCAD) or rare cases with transient neonatal diabetes	Candidate gene	[98, 99]
IER3IP1	Recessive	Permanent neonatal diabetes with microcephaly, simplified gyral pattern and epilepsy	Linkage	[100]

IL2RA	Recessive	Immune dysregulation, polyendocrinopathy, enteropathy (IPEX) syndrome	Candidate gene	[101]
INS	Dominant (often *de novo*) or recessive	Permanent or transient neonatal diabetes or MODY	Linkage	[17, 18, 102]
KCNJ11	Dominant (often *de novo*)	Permanent or transient neonatal diabetes, DEND syndrome or MODY	Candidate gene	[14, 103, 104]
LRBA	Recessive	Permanent neonatal diabetes and additional autoimmune features	Exome sequencing	[42]
MNX1	Recessive	Permanent neonatal diabetes	Candidate gene	[27]
NEUROD1	Dominant or recessive	MODY (dominant) or Permanent neonatal diabetes with neurological abnormalities (recessive)	Candidate gene	[105, 106]
NEUROG3	Recessive	Permanent neonatal diabetes with enteric anendocrinosis	Candidate gene	[107]

NKX2-2	Recessive	Permanent neonatal diabetes and corpus callosum hypoplasia	Candidate gene	[27]
PDX1	Dominant or recessive	MODY (dominant) or Permanent neonatal diabetes (recessive)	Linkage	[22, 108]
PTF1A	Recessive	Permanent neonatal diabetes with cerebellar agenesis/Isolated pancreatic agenesis (enhancer mutations)	Linkage/Genome sequencing	[23, 72]
RFX6	Dominant or recessive	Permanent neonatal diabetes with intestinal atresia and hepatobiliary abnormalities (Recessive) or MODY (dominant)	Linkage/Candidate gene	[33, 34]
SLC2A2	Recessive	Fanconi-Bickel syndrome	Candidate gene	[109]
SLC19A2	Recessive	Thiamine responsive megaloblastic anaemia, diabetes and deafness (TRMA) syndrome	Linkage	[110]
SLC29A3	Recessive	H syndrome	Candidate gene	[111]
STAT3	Dominant	Early-onset polyautoimmunity	Exome sequencing	[41, 55]

		including neonatal/early onset diabetes		
TRMT10A	Recessive	Microcephaly, short stature, and impaired glucose metabolism	Linkage	[24, 25]
WFS1	Dominant or recessive	Wolfram syndrome (recessive), isolated adult onset diabetes (dominant), neonatal/infancy onset diabetes, congenital sensorineural deafness and congenital cataracts (dominant *de novo*)	Linkage/Exome sequencing	[39, 60, 67]
ZBTB20	Dominant (mostly *de novo*)	Primrose syndrome	Exome sequencing	[58]
ZFP57	Recessive	Syndromic transient neonatal diabetes	Linkage	[112]
m.3243A>G (Mitochondrial)	Maternal	Diabetes and deafness	Candidate	[113]

Figure 1: Monogenic diabetes gene in the beta cell. Schematic representation of the monogenic diabetes genes (black font) and their subcellular localisation in the beta cell.

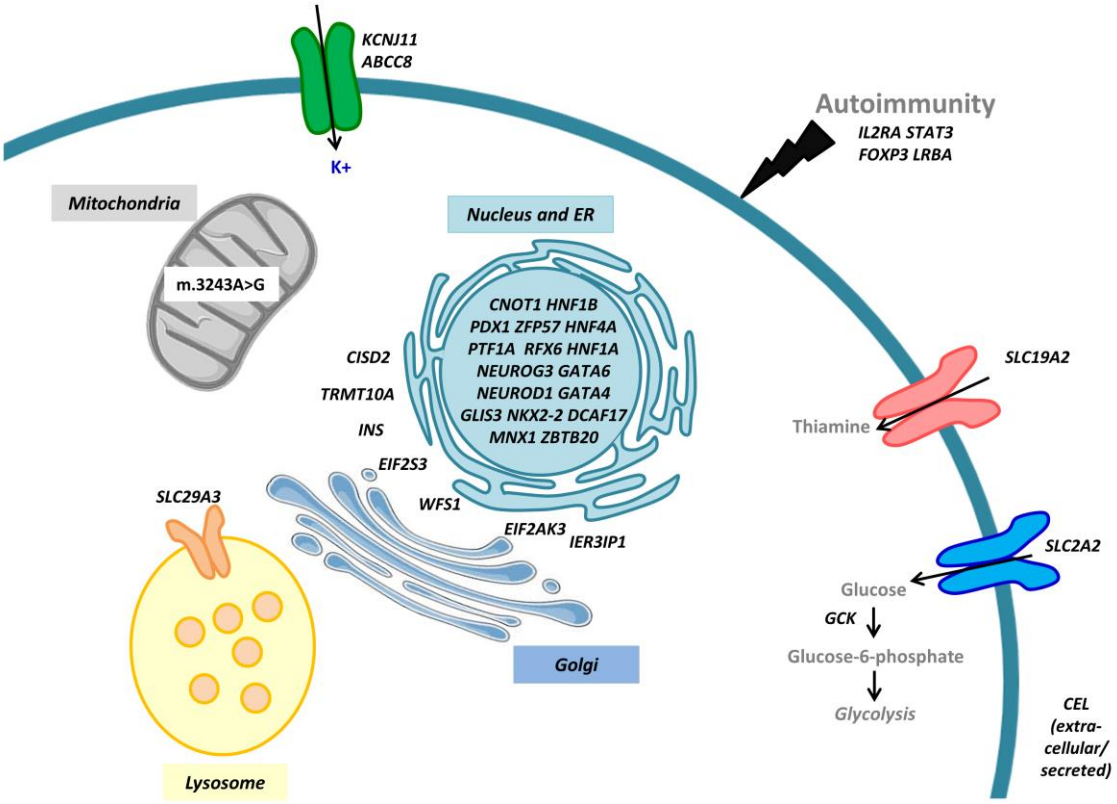

KCNJ11
ABCC8
K+
Autoimmunity
IL2RA STAT3
FOXP3 LRBA
Mitochondria
m.3243A>G
Nucleus and ER
CNOT1 HNF1B
PDX1 ZFP57 HNF4A
PTF1A RFX6 HNF1A
NEUROG3 GATA6
NEUROD1 GATA4
GLIS3 NKX2-2 DCAF17
MNX1 ZBTB20
CISD2
TRMT10A
INS
EIF2S3
WFS1
EIF2AK3
IER3IP1
Golgi
SLC29A3
Lysosome
SLC19A2
Thiamine
SLC2A2
Glucose
GCK
Glucose-6-phosphate
Glycolysis
CEL
(extra-
cellular/
secreted)

Figure 1.

References

[1] Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? Diabetologia. 2010;53:2504-8.
[2] Iafusco D, Stazi MA, Cotichini R, Cotellessa M, Martinucci ME, Mazzella M, et al. Permanent diabetes mellitus in the first year of life. Diabetologia. 2002;45:798-804.
[3] Slingerland AS, Shields BM, Flanagan SE, Bruining GJ, Noordam K, Gach A, et al. Referral rates for diagnostic testing support an incidence of permanent neonatal diabetes in three European countries of at least 1 in 260,000 live births. Diabetologia. 2009;52:1683-5.
[4] Pearson ER, Pruhova S, Tack CJ, Johansen A, Castleden HA, Lumb PJ, et al. Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection. Diabetologia. 2005;48:878-85.
[5] Pearson ER, Flechtner I, Njolstad PR, Malecki MT, Flanagan SE, Larkin B, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med. 2006;355:467-77.
[6] Rafiq M, Flanagan SE, Patch AM, Shields BM, Ellard S, Hattersley AT. Effective treatment with oral sulfonylureas in patients with diabetes due to sulfonylurea receptor 1 (SUR1) mutations. Diabetes Care. 2008;31:204-9.
[7] Shepherd M, Hattersley AT. 'I don't feel like a diabetic any more': the impact of stopping insulin in patients with maturity onset diabetes of the young following genetic testing. Clin Med (Lond). 2004;4:144-7.
[8] Greeley SA, John PM, Winn AN, Ornelas J, Lipton RB, Philipson LH, et al. The cost-effectiveness of personalized genetic medicine: the case of genetic testing in neonatal diabetes. Diabetes Care. 2011;34:622-7.
[9] Mlynarski W, Tarasov AI, Gach A, Girard CA, Pietrzak I, Zubcevic L, et al. Sulfonylurea improves CNS function in a case of intermediate DEND syndrome caused by a mutation in KCNJ11. Nat Clin Pract Neurol. 2007;3:640-5.
[10] Slingerland AS, Nuboer R, Hadders-Algra M, Hattersley AT, Bruining GJ. Improved motor development and good long-term glycaemic control with sulfonylurea treatment in a patient with the syndrome of intermediate developmental delay, early-onset generalised epilepsy and neonatal diabetes associated with the V59M mutation in the KCNJ11 gene. Diabetologia. 2006;49:2559-63.
[11] Slingerland AS, Hurkx W, Noordam K, Flanagan SE, Jukema JW, Meiners LC, et al. Sulphonylurea therapy improves cognition in a patient with the V59M KCNJ11 mutation. Diabet Med. 2008;25:277-81.
[12] De Franco E, Flanagan SE, Houghton JA, Lango Allen H, Mackay DJ, Temple IK, et al. The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. Lancet. 2015;386:957-63.
[13] Millman JR, Pagliuca FW. Autologous Pluripotent Stem Cell-Derived beta-Like Cells for Diabetes Cellular Therapy. Diabetes. 2017;66:1111-20.
[14] Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med. 2004;350:1838-49.
[15] Babenko AP, Polak M, Cave H, Busiah K, Czernichow P, Scharfmann R, et al. Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. N Engl J Med. 2006;355:456-66.
[16] Proks P, Arnold AL, Bruining J, Girard C, Flanagan SE, Larkin B, et al. A heterozygous activating mutation in the sulphonylurea receptor SUR1 (ABCC8) causes neonatal diabetes. Hum Mol Genet. 2006;15:1793-800.
[17] Garin I, Edghill EL, Akerman I, Rubio-Cabezas O, Rica I, Locke JM, et al. Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. Proc Natl Acad Sci U S A. 2010;107:3105-10.

[18] Stoy J, Edghill EL, Flanagan SE, Ye H, Paz VP, Pluzhnikov A, et al. Insulin gene mutations as a cause of permanent neonatal diabetes. Proc Natl Acad Sci U S A. 2007;104:15040-4.
[19] Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). Nature. 1996;384:455-8.
[20] Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature. 1996;384:458-60.
[21] Vionnet N, Stoffel M, Takeda J, Yasuda K, Bell GI, Zouali H, et al. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. Nature. 1992;356:721-2.
[22] Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. Nat Genet. 1997;15:106-10.
[23] Sellick GS, Barker KT, Stolte-Dijkstra I, Fleischmann C, Coleman RJ, Garrett C, et al. Mutations in PTF1A cause pancreatic and cerebellar agenesis. Nat Genet. 2004;36:1301-5.
[24] Igoillo-Esteve M, Genin A, Lambert N, Desir J, Pirson I, Abdulkarim B, et al. tRNA methyltransferase homolog gene TRMT10A mutation in young onset diabetes and primary microcephaly in humans. PLoS Genet. 2013;9:e1003888.
[25] Gillis D, Krishnamohan A, Yaacov B, Shaag A, Jackman JE, Elpeleg O. TRMT10A dysfunction is associated with abnormalities in glucose homeostasis, short stature and microcephaly. J Med Genet. 2014;51:581-6.
[26] Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, et al. High-throughput discovery of novel developmental phenotypes. Nature. 2016;537:508-14.
[27] Flanagan SE, De Franco E, Lango Allen H, Zerah M, Abdul-Rasoul MM, Edge JA, et al. Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man. Cell Metab. 2014;19:146-54.
[28] Sussel L, Kalamaras J, Hartigan-O'Connor DJ, Meneses JJ, Pedersen RA, Rubenstein JL, et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. Development. 1998;125:2213-21.
[29] Briscoe J, Sussel L, Serup P, Hartigan-O'Connor D, Jessell TM, Rubenstein JL, et al. Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature. 1999;398:622-7.
[30] Qi Y, Cai J, Wu Y, Wu R, Lee J, Fu H, et al. Control of oligodendrocyte differentiation by the Nkx2.2 homeodomain transcription factor. Development. 2001;128:2723-33.
[31] Ross AJ, Ruiz-Perez V, Wang Y, Hagan DM, Scherer S, Lynch SA, et al. A homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis. Nat Genet. 1998;20:358-61.
[32] Harrison KA, Thaler J, Pfaff SL, Gu H, Kehrl JH. Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. Nat Genet. 1999;23:71-5.
[33] Patel KA, Kettunen J, Laakso M, Stancakova A, Laver TW, Colclough K, et al. Heterozygous RFX6 protein truncating variants are associated with MODY with reduced penetrance. Nat Commun. 2017;8:888.
[34] Smith SB, Qu HQ, Taleb N, Kishimoto NY, Scheel DW, Lu Y, et al. Rfx6 directs islet formation and insulin production in mice and humans. Nature. 2010;463:775-80.
[35] Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012;20:490-7.
[36] Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J Hum Genet. 2012;57:621-32.
[37] Konrad K, C FL, Grace T, B CB, Jessica A, Qingbo W, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-offunction intolerance across human protein-coding genes. bioRxiv; 2019.

[38] Allen HL, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat Genet. 2011;44:20-2.
[39] De Franco E, Flanagan SE, Yagi T, Abreu D, Mahadevan J, Johnson MB, et al. Dominant ER Stress-Inducing WFS1 Mutations Underlie a Genetic Syndrome of Neonatal/Infancy-Onset Diabetes, Congenital Sensorineural Deafness, and Congenital Cataracts. Diabetes. 2017;66:2044-53.
[40] De Franco E, Watson RA, Weninger WJ, Wong CC, Flanagan SE, Caswell R, et al. A Specific CNOT1 Mutation Results in a Novel Syndrome of Pancreatic Agenesis and Holoprosencephaly through Impaired Pancreatic and Neurological Development. Am J Hum Genet. 2019;104:985-9.
[41] Flanagan SE, Haapaniemi E, Russell MA, Caswell R, Allen HL, De Franco E, et al. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. Nat Genet. 2014;46:812-4.
[42] Johnson MB, De Franco E, Lango Allen H, Al Senani A, Elbarbary N, Siklar Z, et al. Recessively Inherited LRBA Mutations Cause Autoimmunity Presenting as Neonatal Diabetes. Diabetes. 2017;66:2316-22.
[43] Skopkova M, Hennig F, Shin BS, Turner CE, Stanikova D, Brennerova K, et al. EIF2S3 Mutations Associated with Severe X-Linked Intellectual Disability Syndrome MEHMO. Hum Mutat. 2017;38:409-25.
[44] Schwitzgebel VM, Mamin A, Brun T, Ritz-Laser B, Zaiko M, Maret A, et al. Agenesis of human pancreas due to decreased half-life of insulin promoter factor 1. J Clin Endocrinol Metab. 2003;88:4398-406.
[45] Thomas IH, Saini NK, Adhikari A, Lee JM, Kasa-Vubu JZ, Vazquez DM, et al. Neonatal diabetes mellitus with pancreatic agenesis in an infant with homozygous IPF-1 Pro63fsX60 mutation. Pediatr Diabetes. 2009;10:492-6.
[46] Al-Shammari M, Al-Husain M, Al-Kharfy T, Alkuraya FS. A novel PTF1A mutation in a patient with severe pancreatic and cerebellar involvement. Clin Genet. 2011;80:196-8.
[47] Tutak E, Satar M, Yapicioglu H, Altintas A, Narli N, Herguner O, et al. A Turkish newborn infant with cerebellar agenesis/neonatal diabetes mellitus and PTF1A mutation. Genet Couns. 2009;20:147-52.
[48] Morrisey EE, Tang Z, Sigrist K, Lu MM, Jiang F, Ip HS, et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes Dev. 1998;12:3579-90.
[49] Watt AJ, Zhao R, Li J, Duncan SA. Development of the mammalian liver and ventral pancreas is dependent on GATA4. BMC Dev Biol. 2007;7:37.
[50] Kruszka P, Berger SI, Weiss K, Everson JL, Martinez AF, Hong S, et al. A CCR4-NOT Transcription Complex, Subunit 1, CNOT1, Variant Associated with Holoprosencephaly. Am J Hum Genet. 2019;104:990-3.
[51] Winkler GS, Mulder KW, Bardwell VJ, Kalkhoven E, Timmers HT. Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. Embo j. 2006;25:3089-99.
[52] Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, et al. Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. Stem Cells. 2012;30:910-22.
[53] Chatila TA, Blaeser F, Ho N, Lederman HM, Voulgaropoulos C, Helms C, et al. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. J Clin Invest. 2000;106:R75-81.
[54] Holland SM, DeLeo FR, Elloumi HZ, Hsu AP, Uzel G, Brodsky N, et al. STAT3 mutations in the hyper-IgE syndrome. N Engl J Med. 2007;357:1608-19.
[55] Milner JD, Vogel TP, Forbes L, Ma CA, Stray-Pedersen A, Niemela JE, et al. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. Blood. 2015;125:591-9.
[56] Lopez-Herrera G, Tampella G, Pan-Hammarstrom Q, Herholz P, Trujillo-Vargas CM, Phadwal K, et al. Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. Am J Hum Genet. 2012;90:986-1001.

[57] Lo B, Zhang K, Lu W, Zheng L, Zhang Q, Kanellopoulou C, et al. AUTOIMMUNE DISEASE. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. Science. 2015;349:436-40.
[58] Cordeddu V, Redeker B, Stellacci E, Jongejan A, Fragale A, Bradley TE, et al. Mutations in ZBTB20 cause Primrose syndrome. Nat Genet. 2014;46:815-7.
[59] Zhang Y, Xie Z, Zhou L, Li L, Zhang H, Zhou G, et al. The zinc finger protein ZBTB20 regulates transcription of fructose-1,6-bisphosphatase 1 and beta cell function in mice. Gastroenterology. 2012;142:1571-80.e6.
[60] Inoue H, Tanizawa Y, Wasson J, Behn P, Kalidas K, Bernal-Mizrachi E, et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). Nat Genet. 1998;20:143-8.
[61] Bespalova IN, Van Camp G, Bom SJ, Brown DJ, Cryns K, DeWan AT, et al. Mutations in the Wolfram syndrome 1 gene (WFS1) are a common cause of low frequency sensorineural hearing loss. Hum Mol Genet. 2001;10:2501-8.
[62] Lesperance MM, Hall JW, 3rd, San Agustin TB, Leal SM. Mutations in the Wolfram syndrome type 1 gene (WFS1) define a clinical entity of dominant low-frequency sensorineural hearing loss. Arch Otolaryngol Head Neck Surg. 2003;129:411-20.
[63] Berry V, Gregory-Evans C, Emmett W, Waseem N, Raby J, Prescott D, et al. Wolfram gene (WFS1) mutation causes autosomal dominant congenital nuclear cataract in humans. Eur J Hum Genet. 2013;21:1356-60.
[64] Eiberg H, Hansen L, Kjer B, Hansen T, Pedersen O, Bille M, et al. Autosomal dominant optic atrophy associated with hearing impairment and impaired glucose regulation caused by a missense mutation in the WFS1 gene. J Med Genet. 2006;43:435-40.
[65] Hogewind BF, Pennings RJ, Hol FA, Kunst HP, Hoefsloot EH, Cruysberg JR, et al. Autosomal dominant optic neuropathy and sensorineual hearing loss associated with a novel mutation of WFS1. Mol Vis. 2010;16:26-35.
[66] Rendtorff ND, Lodahl M, Boulahbel H, Johansen IR, Pandya A, Welch KO, et al. Identification of p.A684V missense mutation in the WFS1 gene as a frequent cause of autosomal dominant optic atrophy and hearing impairment. Am J Med Genet A. 2011;155a:1298-313.
[67] Bonnycastle LL, Chines PS, Hara T, Huyghe JR, Swift AJ, Heikinheimo P, et al. Autosomal dominant diabetes arising from a Wolfram syndrome 1 mutation. Diabetes. 2013;62:3943-50.
[68] Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511:344-7.
[69] Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med. 2010;362:1181-91.
[70] Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, et al. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Res. 2015;4:1075.
[71] Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13:278-89.
[72] Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, et al. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat Genet. 2014;46:61-4.
[73] Edghill EL, Dix RJ, Flanagan SE, Bingley PJ, Hattersley AT, Ellard S, et al. HLA genotyping supports a nonautoimmune etiology in patients diagnosed with diabetes under the age of 6 months. Diabetes. 2006;55:1895-8.
[74] Johnson MB, Patel KA, De Franco E, Houghton JAL, McDonald TJ, Ellard S, et al. A type 1 diabetes genetic risk score can discriminate monogenic autoimmunity with diabetes from early-onset clustering of polygenic autoimmunity with diabetes. Diabetologia. 2018;61:862-9.
[75] Oram RA, Patel K, Hill A, Shields B, McDonald TJ, Jones A, et al. A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. Diabetes Care. 2016;39:337-44.

[76] Patel KA, Oram RA, Flanagan SE, De Franco E, Colclough K, Shepherd M, et al. Type 1 Diabetes Genetic Risk Score: A Novel Tool to Discriminate Monogenic and Type 1 Diabetes. Diabetes. 2016;65:2094-9.
[77] Klimentidis YC, Wineinger NE, Vazquez AI, de Los Campos G. Multiple metabolic genetic risk scores and type 2 diabetes risk in three racial/ethnic groups. J Clin Endocrinol Metab. 2014;99:E1814-8.
[78] Lall K, Magi R, Morris A, Metspalu A, Fischer K. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. Genet Med. 2017;19:322-9.
[79] Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36:928-30.
[80] Strande NT, Riggs ER, Buchanan AH, Ceyhan-Birsoy O, DiStefano M, Dwight SS, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. Am J Hum Genet. 2017;100:895-906.
[81] Mackay DJ, Boonen SE, Clayton-Smith J, Goodship J, Hahnemann JM, Kant SG, et al. A maternal hypomethylation syndrome presenting as transient neonatal diabetes mellitus. Hum Genet. 2006;120:262-9.
[82] Bowman P, Flanagan SE, Edghill EL, Damhuis A, Shepherd MH, Paisey R, et al. Heterozygous ABCC8 mutations are a cause of MODY. Diabetologia. 2012;55:123-7.
[83] Ellard S, Flanagan SE, Girard CA, Patch AM, Harries LW, Parrish A, et al. Permanent neonatal diabetes caused by dominant, recessive, or compound heterozygous SUR1 mutations with opposite functional effects. Am J Hum Genet. 2007;81:375-82.
[84] Raeder H, Johansson S, Holm PI, Haldorsen IS, Mas E, Sbarra V, et al. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. Nat Genet. 2006;38:54-62.
[85] Amr S, Heisey C, Zhang M, Xia XJ, Shows KH, Ajlouni K, et al. A homozygous mutation in a novel zinc-finger protein, ERIS, is responsible for Wolfram syndrome 2. Am J Hum Genet. 2007;81:673-83.
[86] Alazami AM, Al-Saif A, Al-Semari A, Bohlega S, Zlitni S, Alzahrani F, et al. Mutations in C2orf37, encoding a nucleolar protein, cause hypogonadism, alopecia, diabetes mellitus, mental retardation, and extrapyramidal syndrome. Am J Hum Genet. 2008;83:684-91.
[87] Alazami AM, Schneider SA, Bonneau D, Pasquier L, Carecchio M, Kojovic M, et al. C2orf37 mutational spectrum in Woodhouse-Sakati syndrome patients. Clin Genet. 2010;78:585-90.
[88] Delepine M, Nicolino M, Barrett T, Golamaully M, Lathrop GM, Julier C. EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. Nat Genet. 2000;25:406-9.
[89] Moortgat S, Desir J, Benoit V, Boulanger S, Pendeville H, Nassogne MC, et al. Two novel EIF2S3 mutations associated with syndromic intellectual disability with severe microcephaly, growth retardation, and epilepsy. Am J Med Genet A. 2016;170:2927-33.
[90] Stanik J, Skopkova M, Stanikova D, Brennerova K, Barak L, Ticha L, et al. Neonatal hypoglycemia, early-onset diabetes and hypopituitarism due to the mutation in EIF2S3 gene causing MEHMO syndrome. Physiol Res. 2018;67:331-7.
[91] Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, Whitesell L, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet. 2001;27:20-1.
[92] Wildin RS, Ramsdell F, Peake J, Faravelli F, Casanova JL, Buist N, et al. X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet. 2001;27:18-20.
[93] D'Amato E, Giacopelli F, Giannattasio A, D'Annunzio G, Bocciardi R, Musso M, et al. Genetic investigation in an Italian child with an unusual association of atrial septal defect, attributable to a new familial GATA4 gene mutation, and neonatal diabetes due to pancreatic agenesis. Diabet Med. 2010;27:1195-200.
[94] Shaw-Smith C, De Franco E, Lango Allen H, Batlle M, Flanagan SE, Borowiec M, et al. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. Diabetes. 2014;63:2888-94.

[95] Catli G, Abaci A, Flanagan SE, De Franco E, Ellard S, Hattersley A, et al. A novel GATA6 mutation leading to congenital heart defects and permanent neonatal diabetes: a case report. Diabetes Metab. 2013;39:370-4.
[96] Njolstad PR, Sovik O, Cuesta-Munoz A, Bjorkhaug L, Massa O, Barbetti F, et al. Neonatal diabetes mellitus due to complete glucokinase deficiency. N Engl J Med. 2001;344:1588-92.
[97] Senee V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, et al. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet. United States2006. p. 682-7.
[98] Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, et al. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet. 1997;17:384-5.
[99] Yorifuji T, Kurokawa K, Mamada M, Imai T, Kawai M, Nishi Y, et al. Neonatal diabetes mellitus and neonatal polycystic, dysplastic kidneys: Phenotypically discordant recurrence of a mutation in the hepatocyte nuclear factor-1beta gene due to germline mosaicism. J Clin Endocrinol Metab. 2004;89:2905-8.
[100] Poulton CJ, Schot R, Kia SK, Jones M, Verheijen FW, Venselaar H, et al. Microcephaly with simplified gyration, epilepsy, and infantile diabetes linked to inappropriate apoptosis of neural progenitors. Am J Hum Genet. 2011;89:265-76.
[101] Caudy AA, Reddy ST, Chatila T, Atkinson JP, Verbsky JW. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like syndrome, and defective IL-10 expression from CD4 lymphocytes. J Allergy Clin Immunol. 2007;119:482-7.
[102] Molven A, Ringdal M, Nordbo AM, Raeder H, Stoy J, Lipkind GM, et al. Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes. Diabetes. 2008;57:1131-5.
[103] Gloyn AL, Reimann F, Girard C, Edghill EL, Proks P, Pearson ER, et al. Relapsing diabetes can result from moderately activating mutations in KCNJ11. Hum Mol Genet. 2005;14:925-34.
[104] Yorifuji T, Nagashima K, Kurokawa K, Kawai M, Oishi M, Akazawa Y, et al. The C42R mutation in the Kir6.2 (KCNJ11) gene as a cause of transient neonatal diabetes, childhood diabetes, or later-onset, apparently type 2 diabetes mellitus. J Clin Endocrinol Metab. 2005;90:3174-8.
[105] Malecki MT, Jhala US, Antonellis A, Fields L, Doria A, Orban T, et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. Nat Genet. 1999;23:323-8.
[106] Rubio-Cabezas O, Minton JA, Kantor I, Williams D, Ellard S, Hattersley AT. Homozygous mutations in NEUROD1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities. Diabetes. 2010;59:2326-31.
[107] Rubio-Cabezas O, Jensen JN, Hodgson MI, Codner E, Ellard S, Serup P, et al. Permanent Neonatal Diabetes and Enteric Anendocrinosis Associated With Biallelic Mutations in NEUROG3. Diabetes. 2011;60:1349-53.
[108] Stoffers DA, Ferrer J, Clarke WL, Habener JF. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet. 1997;17:138-9.
[109] Santer R, Groth S, Kinner M, Dombrowski A, Berry GT, Brodehl J, et al. The mutation spectrum of the facilitative glucose transporter gene SLC2A2 (GLUT2) in patients with Fanconi-Bickel syndrome. Hum Genet. 2002;110:21-9.
[110] Labay V, Raz T, Baron D, Mandel H, Williams H, Barrett T, et al. Mutations in SLC19A2 cause thiamine-responsive megaloblastic anaemia associated with diabetes mellitus and deafness. Nat Genet. 1999;22:300-4.
[111] Cliffe ST, Kramer JM, Hussain K, Robben JH, de Jong EK, de Brouwer AP, et al. SLC29A3 gene is mutated in pigmented hypertrichosis with insulin-dependent diabetes mellitus syndrome and interacts with the insulin signaling pathway. Hum Mol Genet. 2009;18:2257-65.
[112] Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet. 2008;40:949-51.

[113] van den Ouweland JM, Lemkes HH, Ruitenbeek W, Sandkuijl LA, de Vijlder MF, Struyvenberg PA, et al. Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. Nat Genet. 1992;1:368-71.